Appl. Comput. Harmon. Anal. 40 (2016) 417–429

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

Low rank matrix completion by alternating steepest descent methods

Jared Tanner [a], Ke Wei [b,*]

[a] *Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK*
[b] *Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, People's Republic of China*

ARTICLE INFO

Article history:
Received 7 April 2015
Received in revised form 29 July 2015
Accepted 4 August 2015
Available online 12 August 2015
Communicated by Charles K. Chui

MSC:
15A29
41A29
65F10
65J20
68Q25
90C26

Keywords:
Matrix completion
Alternating minimization
Gradient descent
Exact line-search

ABSTRACT

Matrix completion involves recovering a matrix from a subset of its entries by utilizing interdependency between the entries, typically through low rank structure. Despite matrix completion requiring the global solution of a non-convex objective, there are many computationally efficient algorithms which are effective for a broad class of matrices. In this paper, we introduce an alternating steepest descent algorithm (ASD) and a scaled variant, ScaledASD, for the fixed-rank matrix completion problem. Empirical evaluation of ASD and ScaledASD on both image inpainting and random problems show they are competitive with other state-of-the-art matrix completion algorithms in terms of recoverable rank and overall computational time. In particular, their low per iteration computational complexity makes ASD and ScaledASD efficient for large size problems, especially when computing the solutions to moderate accuracy such as in the presence of model misfit, noise, and/or as an initialization strategy for higher order methods. A preliminary convergence analysis is also presented.

1. Introduction

The problem of recovering a low rank matrix from partial entries – also known as matrix completion – arises in a wide variety of practical contexts, such as model reduction [17], pattern recognition [8], and machine learning [1,2]. From the pioneering work on low rank approximation by Fazel [10] and matrix completion by Candès and Recht [6], this problem has received intensive investigations both from theoretical and algorithmic aspects, see [3–5,7,12,14,16,18–21,23–25] and references therein for a partial review. These

* Corresponding author.
E-mail addresses: tanner@maths.ox.ac.uk (J. Tanner), makwei@ust.hk (K. Wei).

http://dx.doi.org/10.1016/j.acha.2015.08.003

matrix completion and low rank approximation techniques rely on the dependencies between entries imposed by the low rank structure. Explicitly seeking the lowest rank matrix consistent with the known entries is expressed as

$$\min_{Z\in\mathbb{R}^{m\times n}} \operatorname{rank}(Z), \text{ subject to } P_\Omega(Z) = P_\Omega(Z^0), \tag{1}$$

where $Z^0 \in \mathbb{R}^{m\times n}$ is the underlying matrix to be reconstructed, Ω is a subset of indices for the known entries, and P_Ω is the associated sampling operator which acquires only the entries indexed by Ω. Problem (1) is non-convex and generally NP-hard [13] due to the rank objective. One of the most widely studied approaches is to replace the rank objective in (1) with its convex relaxation, the Schatten 1-norm (also known as nuclear norm) which is the sum of the singular values, i.e. $\min_{Z\in\mathbb{R}^{m\times n}} \|Z\|_*$, subject to $P_\Omega(Z) = P_\Omega(Z^0)$. It has been proven that, provided the singular vectors are weakly correlated with the canonical basis, the solution of (1) can be obtained by solving the aforementioned convex relaxation [6]. Alternative to the convex relaxation, there have been many algorithms which are designed to attempt to solve for the global minima of (1) directly; many of them are adaptations of algorithms for compressed sensing, such as the hard thresholding algorithms [3,14,16,21]. The iterative thresholding algorithms typically update a current rank r estimate along a line-search direction which departs the manifold of rank r matrices, and then projects back to the rank r manifold by computing the nearest matrix in Frobenius norm. The most direct implementation of these algorithms require computing a partial singular value decomposition (SVD) in each iteration. The computational complexity of computing the SVD has complexity of $O(n^3)$ when r, m and n are proportional, causing computing the SVD to be the dominant computational cost per iteration and limits their applicability for large n. A subclass of methods which further exploit the manifold structure in their line-search updates achieve superior efficiency, for $r \ll n$, of order $O(r^3)$. Examples include LRGeomCG [23] which is a nonlinear conjugate gradient method, ScGrassMC [20] which uses a scaled gradient in the subspace update, and the algorithms presented in [18,19] which use other geometries and metrics.

To circumvent the high computational cost of an SVD, other algorithms explicitly remain on the manifold of rank r matrices by using the factorization $Z = XY$ where $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{r\times n}$. Based on this simple factorization model, rather than solving (1), algorithms are designed to solve the non-convex problem $\min_{X,Y} f(X,Y)$ where

$$f(X,Y) := \frac{1}{2}\|P_\Omega(Z^0) - P_\Omega(XY)\|_F^2. \tag{2}$$

Algorithms for the solution of (2) usually follow an alternating minimization scheme, with PowerFactorization [12] and LMaFit [24] two representatives. We present alternating steepest descent (ASD), and a scaled variant ScaledASD, which incorporate an exact line-search to update the solutions for the model (2). *In so doing ASD and ScaledASD have a lower per iteration complexity than PowerFactorization, allowing a more efficient exploration of the manifold in the early iterations where the current estimate is inaccurate. Moreover, ASD and ScaledASD are able to recover matrices of substantially higher rank than can LMaFit.*

The manuscript is outlined as follows. In Section 2 we briefly review the alternating minimization algorithms PowerFactorization and LMaFit which motivate ASD. In Section 3 we propose ASD for (2), which replaces the least square subproblem solution in PowerFactorization with exact line-search so as to reduce the per iteration computational cost. In Section 4 we present ScaledASD, a version of ASD which is accelerated by scaling the search directions adaptively to improve the asymptotic convergence rate. A preliminary convergence analysis for ASD and ScaledASD are given in Section 3 and Section 4 respectively. Numerical experiments presented in Section 5 contrast the aforementioned algorithms and show that ScaledASD is highly efficient, particularly for large problems when solved to moderate accuracy.